The fractality of polar codes

Conference Paper

Author(s):
Geiger, Bernhard C.

Publication date:
2016

Permanent link:
https://doi.org/10.3929/ethz-a-010645921

The Fractality of Polar Codes

Bernhard C. Geiger
Institute for Communications Engineering, Technische Universität München, Germany
geiger@ieee.org

***Abstract*—The generator matrix of a polar code is obtained by selecting rows from the Kronecker product of a lower-triangular binary square matrix. The selection is based on the Bhattacharyya parameter of the row, which is closely related to the error probability of the corresponding input bit under sequential decoding. This work investigates the properties of the index set pointing to those rows in the infinite blocklength limit. In particular, the Lebesgue measure, the Hausdorff dimension, and the self-similarity of this set will be discussed. It is shown that these index sets fulfill several properties that are common to fractals.**

I. Introduction

Applying the polarization transform proposed by Arıkan [1] to sufficiently many instances of a binary-input memoryless channel, causes a portion of the resulting channels to have a capacity close to one, while the remaining portion has a capacity close to zero. These *polarized channels* can thus be split into two sets: The set of "good" channels, and the set of "bad" channels. Despite their importance for code construction, very little is known about their structure. A recent exception is the work by Renes, Sutter, and Hassani, stating conditions under which polarized sets are aligned, i.e., under which the good (bad) channels derived from one binary-input memoryless channel are a subset of the good (bad) channels derived from another [2].

Polar codes are Kronecker product-based codes. Such a code of block-length 2^n is based on the n-fold Kronecker product $G(n) := F^{\otimes n}$, where

$$F := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \quad (1)$$

Following the terminology of [3], a rate-$K/2^n$ Kronecker product-based code is uniquely defined by a set $\mathcal{F}$ of K indices: Its generator matrix is the submatrix of $G(n)$ consisting of the rows indexed by $\mathcal{F}$. For polar codes, in which each row of $G(n)$ can be interpreted as a (partially polarized) channel, $\mathcal{F}$ consists of rows corresponding to the K channels with the lowest Bhattacharyya parameter [4] (see Section II).

That Kronecker product-based codes, such as polar codes [1] or Reed-Muller codes, possess a fractal nature has been observed in [3], where it was noted that $G(n)$ resembles a Sierpinski triangle. Much earlier, Abbe suspected that the set of "good" channels has fractal nature [5]. Nevertheless, to the best of the author's knowledge, no definite statement regarding this fractal nature has been made yet. In this paper, we try to fill this gap and present results about the set of "good" channels (Sections III). Specifically, we study the properties of the set $\mathcal{F}$ for infinite blocklengths, i.e., for $n \to \infty$.

To simplify analysis, we represent every infinite binary sequence indexed in $\mathcal{F}$ by a point in the unit interval $[0, 1]$. Let $\Omega = \{0, 1\}^\infty$ be the set of infinite binary sequences, and let $b := (b_1 b_2 \cdots) \in \Omega$ be an arbitrary such sequence. We abbreviate $b^n := (b_1 b_2 \cdots b_n)$. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space with $\mathfrak{B}$ the Borel field generated by the cylinder sets $S(b^n) := \{w \in \Omega\colon w_1 = b_1, \dots, w_n = b_2\}$ and $\mathbb{P}$ a probability measure satisfying $\mathbb{P}(S(b^n)) = 1/2^n$. The following function $f\colon \Omega \to [0, 1]$ permits us to convert these sequences to real numbers:

$$f(b) := \sum_{n=1}^{\infty} \frac{b_n}{2^n} \quad (2)$$

Letting $\mathbb{D} := [0, 1] \cap \{p/2^n\colon p \in \mathbb{Z}, n \in \mathbb{N}\}$ denote the set of dyadic rationals in the unit interval, we recognize that f is not injective:

Example 1. f maps both $b = (01111111 \cdots)$ and $b = (10000000 \cdots)$ to 0.5. We call the latter binary expansion *terminating*.

However, as the following lemma shows, f is bijective if we exclude the dyadic rationals:

Lemma 1 ([6, Exercises 7-10, p. 80])**.** *Let $\mathfrak{B}_{[0,1]}$ be the Borel σ-algebra on $[0, 1]$ and let λ be the Lebesgue measure. Then, the function f in* (2) *satisfies the following properties:*
1) *f is measurable w.r.t. $\mathfrak{B}_{[0,1]}$*
2) *f is bijective on $\Omega \setminus f^{-1}(\mathbb{D})$*
3) *for all $I \in \mathfrak{B}_{[0,1]}$, $\mathbb{P}(f^{-1}(I)) = \lambda(I)$*

We believe that the results we prove in the following not only improve our understanding of polar codes: Since its introduction in 2009, the polarization technique proposed by Arıkan has found its way into areas different from polar coding. Haghighatshoar and Abbe showed in the context of compression of analog sources that Rényi information dimension can be polarized [7], and Abbe and Wigderson used polarization for the construction of high-girth matrices [8]. Recently, Nasser proved that a binary operation is polarizing if and only if it is uniformity preserving and its inverse is strongly ergodic [9], [10]. We believe that our results might carry over to these areas as well and point to possible extensions in Section IV.

II. Preliminaries for Polar Codes

We adopt the notation of [1]: Let $W\colon \{0, 1\} \to \mathcal{Y}$ be a binary-input memoryless channel with output alphabet $\mathcal{Y}$, capacity $0 < I(W) < 1$, and with Bhattacharyya parameter

$$Z(W) := \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}. \quad (3)$$

That $Z(W)=0 \Leftrightarrow I(W)=1$ and $Z(W)=1 \Leftrightarrow I(W)=0$ is a direct consequence of [1, Prop. 1].

The heart of Arıkan's polarization technique is that two channel uses of W can be *combined and split* into one use of a "worse" channel

$$W_2^0(y_1^2|u_1) := \frac{1}{2}\sum_{u_2} W(y_1|u_1 \oplus u_2)W(y_2|u_2) \tag{4a}$$

and one use of a better channel

$$W_2^1(y_1^2, u_1|u_2) := \frac{1}{2} W(y_1|u_1 \oplus u_2)W(y_2|u_2) \tag{4b}$$

where $u_1, u_2 \in \{0,1\}$ and $y_1, y_2 \in \mathcal{Y}$. In essence, the combining operation codes two input bits by F in (1) and transmits the coded bits over W via two channel uses, creating a vector channel. The splitting operation splits this vector channel into the two binary-input memoryless channels indicated in (4). Of these, the better (worse) channel has a strictly larger (smaller) capacity than the original channel W, i.e., $I(W_2^0) < I(W) < I(W_2^1)$, while the sum capacity equals twice the capacity of the original channel, i.e., $I(W_2^0) + I(W_2^1) = 2I(W)$ [1, Prop. 4].

The effect of combining and splitting on the channel capacities $I(W_2^0)$ and $I(W_2^1)$ admits no closed-form expression; the effect on the Bhattacharyya parameter at least admits bounds:

Lemma 2 ([1, Prop. 5 & 7])**.**

$$Z(W_2^1) = g_1(Z(W)) = Z^2(W) < Z(W) \tag{5a}$$

$$Z(W) < Z(W_2^0) \le g_0(Z(W)) = 2Z(W) - Z^2(W) \tag{5b}$$

with equality if W is a binary erasure channel.

Channels with larger blocklengths 2^n, $n > 1$, can either be obtained by direct n-fold combining (using the matrix $G(n)$) and n-fold splitting, or by recursive pairwise combining and splitting. For $b^n \in \{0,1\}^n$, we obtain

$$\left(W_{2^n}^{b^n}, W_{2^n}^{b^n}\right) \to \left(W_{2^{n+1}}^{b^n 0}, W_{2^{n+1}}^{b^n 1}\right) \tag{6}$$

where $b^n 0$ and $b^n 1$ denote the sequences of zeros and ones obtained by appending 0 and 1 to b^n, respectively. Note that g_1 and g_0 from Lemma 2 are non-negative and non-decreasing functions mapping the unit interval onto itself, hence the inequality in (5b) is preserved under composition:

$$Z(W_{2^n}^{b^n}) \le p_{b^n}(Z(W)) := g_{b_n}(g_{b_{n-1}}(\cdots g_{b_1}(Z(W))\cdots)) \tag{7}$$

The channel polarization theorem shows that, with probability one, after infinitely many combinations and splits, only perfect or useless channels remain, i.e., either $I(W_\infty^b) = 1$ or $I(W_\infty^b) = 0$ for $b \in \{0,1\}^\infty$. This is made precise in:

Proposition 1 ([1, Prop. 10])**.** *With probability one, the limit RV $I_\infty(b) := I(W_\infty^b)$ takes values in the set $\{0,1\}$: $\mathbb{P}(I_\infty = 1) = I(W)$ and $\mathbb{P}(I_\infty = 0) = 1 - I(W)$.*

If the polarization procedure is stopped at a finite blocklength 2^n for n large enough, it can still be shown that the vast majority of the resulting 2^n channels are either almost perfect or almost useless, in the sense that the channel capacities are close to one or to zero (or, that the corresponding Bhattacharyya parameters are close to zero or to one). The idea of polar coding is to transmit data only on those channels that are almost perfect: n-fold combining, which employs the matrix $G(n)$, leads to 2^n virtual channels, each corresponding to a row of $G(n)$. The channels with high capacity are indicated by the set $\mathcal{F}$, and the generator matrix of the corresponding polar code is precisely the submatrix of $G(n)$ consisting of those indicated rows.

The difficulty of polar coding lies in code construction, i.e., in determining *which* channels/row indices are in the set $\mathcal{F}$. This immediately translates to the question which sequences $b \in \{0,1\}^\infty$ correspond to combinations and splits leading to a perfect channel (or which finite-length sequences b^n lead to channels with capacity sufficiently close to one). Determining the capacity of the virtual channels is an inherently difficult operation, since, whenever W is not a binary erasure channel (BEC), the cardinality of the output alphabet increases exponentially in 2^n [11, Ch. 3.3], [12, p. 36]. To circumvent this problem, Tal and Vardy presented an approximate construction method in [13], that relies on working with reduced output alphabet channels that are either upgraded or degraded w.r.t. the real channel. As these upgrading/degrading properties – mentioned earlier in Korada's PhD thesis [12] – will play a fundamental role in this work, we present

Definition 1 (Channel Up- and Degrading)**.** A channel $W^-\colon \{0,1\} \to \mathcal{Z}$ is *degraded* w.r.t. the channel W (short: $W^- \preccurlyeq W$) if there exists a channel $P\colon \mathcal{Y} \to \mathcal{Z}$ such that

$$W^-(z|u) = \sum_{y\in\mathcal{Y}} W(y|u)P(z|y). \tag{8}$$

A channel $W^+\colon \{0,1\} \to \mathcal{Z}$ is *upgraded* w.r.t. the channel W (short: $W^+ \succcurlyeq W$) if there exists a channel $P\colon \mathcal{Z} \to \mathcal{Y}$ such that

$$W(y|u) = \sum_{z\in\mathcal{Z}} W^+(z|u)P(y|z). \tag{9}$$

Moreover, $W^+ \succcurlyeq W$ if and only if $W \preccurlyeq W^+$.

The upgraded (degraded) approximation remains upgraded (degraded) during combining and splitting:

Lemma 3 ([12, Lem. 4.7] & [13, Lem. 3])**.** *Assume that $W^- \preccurlyeq W \preccurlyeq W^+$. Then,*

$$I(W^-) \le I(W) \le I(W^+) \tag{10a}$$

$$Z(W^-) \ge Z(W) \ge Z(W^+) \tag{10b}$$

$$(W^-)_2^1 \preccurlyeq W_2^1 \preccurlyeq (W^+)_2^1 \tag{10c}$$

$$(W^-)_2^0 \preccurlyeq W_2^0 \preccurlyeq (W^+)_2^0. \tag{10d}$$

It can be shown that the better channel (4b) obtained from combining and splitting is upgraded w.r.t. the original channel (as already mentioned in [11, p. 9]). That the worse channel (4a) is degraded holds at least for the BEC:

Lemma 4. *$W \preccurlyeq W_2^1$. If W is a BEC, then $W_2^0 \preccurlyeq W \preccurlyeq W_2^1$.*

Proof. The proof of the first part follows by choosing

$$P(y|y_1^2, u_1) = \begin{cases} 1, & \text{if } y = y_2 \\ 0, & \text{else.} \end{cases} \tag{11}$$

For the BEC, note that if W has erasure probability ϵ, then W_2^1 is a BEC with erasure probability ϵ^2 and W_2^0 is a BEC with erasure probability $2\epsilon - \epsilon^2$ [1, Prop. 6]. The channel W_2^1 is an upgrade of W, because it can be degraded to W by appending a BEC with erasure probability $\epsilon/(1+\epsilon)$. The channel W_2^0 is degraded w.r.t. W by appending a BEC with erasure probability ϵ. □

III. Properties of the Sets $\mathcal{G}$ and $\mathcal{B}$

In this section we develop the properties of the sets of good and bad channels. For the sake of brevity, we only sketch the proofs here; complete proofs are given in [14].

Definition 2 (The Good and the Bad Channels)**.** Let $\mathcal{G}$ denote the set of good channels, i.e.,

$$x \in \mathcal{G} \Leftrightarrow \exists b \in f^{-1}(x)\colon I(W_\infty^b) = 1; \tag{12}$$

let $\mathcal{B}$ denote the set of bad channels, i.e.,

$$x \in \mathcal{B} \Leftrightarrow \exists b \in f^{-1}(x)\colon I(W_\infty^b) = 0. \tag{13}$$

Proposition 2. *For almost all x, there exists a value $0 \leq \vartheta(x) \leq 1$ such that $Z(W) < \vartheta(x)$ implies $x \in \mathcal{G}$. If W is a BEC, then additionally $Z(W) > \vartheta(x)$ implies $x \in \mathcal{B}$.*

Sketch of Proof: This proposition is an adaption of [15, Lem. 11] to our setting: The lemma states that, for $\mathbb{P}$-almost every sequence b, there is a threshold $\theta(b)$ such that $\lim_{n\to\infty} p_{b^n}(z)$ converges to zero (one) if z is smaller (larger) than $\theta(b)$. The rest follows from Lemma 2. ■

Note that if W is not a BEC, it may occur that $Z(W) > \vartheta(f(b))$ while still $I(W_\infty^b) = 1$. This in turn opens the question whether the set of good channels is (almost surely) increasing with decreasing Bhattacharyya parameter: Are there channels W and W' (from the same family) with good channel sets $\mathcal{G}$ and $\mathcal{G}'$, respectively, such that $Z(W) > Z(W') > \vartheta(f(b))$, but $I(W_\infty^b) = 1$ and $I({W'}_\infty^b) = 0$? We leave this question for future research but mention that Proposition 2 answers it negatively for BECs: The set of good channels for a BEC is also good for any binary-input memoryless channel with a smaller Bhattacharyya parameter [16].

Example 2. For $x \in \mathbb{D}$, $\vartheta(x) = 1$: If $Z(W) < 1$, i.e., if the channel is not completely useless a priori, the non-terminating expansion of x will make it a perfect channel (cf. Proposition 3).

Example 3. Let $x = 2/3$, hence $f^{-1}(x) = 101010101\cdots$. The binary expansion is recurring. It thus suffices to consider exactly one period of the recurring sequence and determine its fixed points. In this case we get $p_{10}(z) = 2z^2 - z^4$. Its fixed point lies at the intersection of $p_{10}(z)$ and z; removing the trivial intersections at $z = 0$ and $z = 1$ leaves two further roots at $(\pm\sqrt{5} - 1)/2$. One of these roots lies outside $[0, 1]$ and is hence irrelevant. The remaining root determines the threshold: $\vartheta(2/3) = (\sqrt{5} - 1)/2$. Now let W be a BEC with erasure probability $\epsilon = Z(W) = \vartheta(2/3)$. Since $\epsilon = \vartheta(2/3)$ is a fixed point of the iterated function system corresponding to the recurring binary expansion, one gets $Z(W_\infty^{f^{-1}(2/3)}) = \epsilon \notin \{0, 1\}$. This example illustrates why Proposition 1 holds only almost surely.

Proposition 3. $\mathcal{G} \cap \mathcal{B} = \mathbb{D}$.

Sketch of Proof: The proof is based on the fact that dyadic rationals admit two possible binary expansions (see Example 1): The Bhattacharyya parameter of the non-terminating expansion $a^k 111\cdots$, for $a^k \in \{0,1\}^k$ an appropriate prefix, is driven down to zero by squaring $Z(W_{2^k}^{a^k})$ infinitely often.

The terminating expansion has the same prefix a^k with the last bit inverted. All binary sequences starting with this prefix lead to a channel that is upgraded w.r.t. the one corresponding to the terminating expansion (Lemmas 3 and 4). By Proposition 1, some sequences with this prefix lead to bad channels, hence the terminating expansion must lead to a bad channel as well. ■

That the intersection of the sets of good and bad channels is non-empty is a direct consequence of the non-injectivity of f. Note further that this intersection cannot be larger, since $\mathbb{D}$ is the only set to which f maps non-injectively. Since $\mathbb{D}$, a common subset of $\mathcal{G}$ and $\mathcal{B}$, is dense in $[0, 1]$, both the set of good channels and the set of bad channels are dense in the unit interval. But even if dyadic rationals are excluded, results about denseness can be proved:

Proposition 4. *$\mathcal{G} \setminus \mathbb{D}$ is dense in $[0, 1]$. If W is a BEC, then also $\mathcal{B} \setminus \mathbb{D}$ is dense in $[0, 1]$.*

Sketch of Proof: We sketch only the first part of the proof, the second part involving BECs follows along the same lines. The proof is based on the polynomial $p_b(z)$. Let b^n be an arbitrary prefix (corresponding to a dyadic rational), leading to a Bhattacharyya parameter $Z(W_{2^n}^{b^n})$. There exists a sequence a^k with one zero and sufficiently many ones such that $p_{a^k}(z) < z$ for all z below a certain threshold $z^*(a^k) > Z(W_{2^n}^{b^n})$. It follows by Lemma 2 that $Z(W_\infty^{b^n a^k a^k \cdots}) \leq p_{a^k a^k \cdots}(Z(W_{2^n}^{b^n})) \to 0$, hence $f(b^n a^k a^k \cdots) \in \mathcal{G}$. Finally, between any two dyadic rationals a rational can be found with binary expansion $b^n a^k a^k \cdots$ that satisfies these properties. This proves that the good channels are dense even excluding the dyadic rationals. The inequality in Lemma 2 is the reason why denseness of bad channels can only be proved for BECs. ■

The proposition states that, at least for the BEC, there is no interval which contains only good channels. Hence, given a specific channel $W_{2^n}^{b^n}$, it is not possible to assume that a well-specified subset of channels (e.g., all $W_\infty^{b^n a}$ for a starting with 1) generated from this channel by combining and splitting will be perfect. The construction algorithm for an infinite-blocklength, *vanishing-error* polar code hence cannot stop at a finite blocklength, as it can be done for a finite-blocklength polar code, cf. [17].

Proposition 5. *$\mathcal{G}$ is Lebesgue measurable and has Lebesgue measure $\lambda(\mathcal{G}) = I(W)$. $\mathcal{B}$ is Lebesgue measurable and has Lebesgue measure $\lambda(\mathcal{B}) = 1 - I(W)$. The Hausdorff dimensions of $\mathcal{G}$ and $\mathcal{B}$ satisfy $d(\mathcal{G}) = 1$ and $d(\mathcal{B}) = 1$.*

Sketch of Proof: The proof for the good channels follows

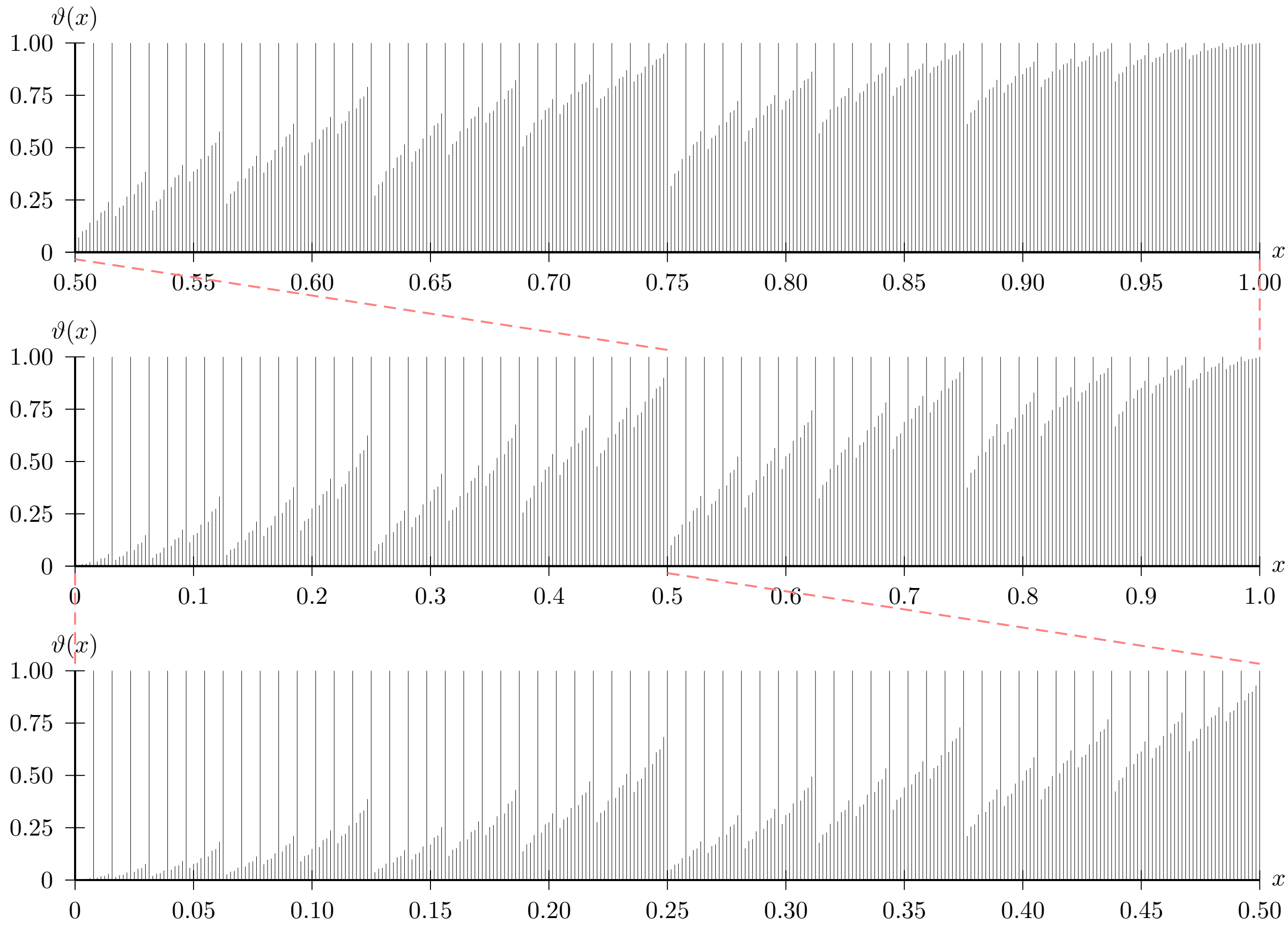

Fig. 1. The polar fractal for a BEC. The center plot shows the thresholds $\vartheta(x)$ for $x \in [0,1]$, while the bottom and the top plots show these thresholds for the scaled and shifted sets $[0,0.5]$ and $[0.5,1]$, respectively. Hence, the thresholds in the top plot are larger than the thresholds in the center plot, which are larger than those in the bottom plot. The set $\mathcal{G}$ is obtained by setting each value in the plot to one (zero) if the erasure probability ϵ is smaller (larger) than the threshold.

from the fact that $\lambda(\mathcal{G}) = \lambda(\mathcal{G} \setminus \mathbb{D})$, from Definition 2 stating

$$x \notin \mathbb{D}\colon x \in \mathcal{G} \Leftrightarrow I(W_\infty^{f^{-1}(x)}) = 1, \tag{14}$$

and from Proposition 1; the proof for the bad channels follows along the same lines. That the Hausdorff dimension of both sets is unity follows from the fact that the one-dimensional Hausdorff measure of a set equals its Lebesgue measure up to a constant [18, eq. (3.4), p. 45]. ■

Note that despite the fact that $\lambda(\mathcal{G} \cup \mathcal{B}) = 1$, $\mathcal{G} \cup \mathcal{B} \subset [0,1]$. The reason is that convergence to good or bad channels is only almost sure, and that there may be channels W_∞^b which are neither good nor bad (see Example 3).

We finally come to the claim that polar codes are fractal. Following Falconer's definition [18, p. xxviii], a set is fractal if it is (at least approximately) self-similar and has detail on arbitrarily small scales, or if its fractal dimension (e.g., its Hausdorff dimension) is larger than its topological dimension. Whether or not the result shown below will convince the reader of this property is a mere question of definition; strictly speaking, we can show only *quasi self-similarity* of $\mathcal{G}$:

Proposition 6. *Let* $\mathcal{G}_n(k) := \mathcal{G} \cap [(k-1)2^{-n}, k2^{-n}]$ *for* $k = 1, \dots, 2^n$. $\mathcal{G} = \mathcal{G}_0(1)$ *is* quasi self-similar *in the sense that, for all* n *and all* k, $\mathcal{G}_n(k) = \mathcal{G}_{n+1}(2k-1) \cup \mathcal{G}_{n+1}(2k)$ *is quasi self-similar to its right half:*

$$\mathcal{G}_n(k) \subset 2\mathcal{G}_{n+1}(2k) - k2^{-n} \tag{15}$$

If W *is a BEC,* $\mathcal{G}_n(k)$ *is quasi self-similar:*

$$2\mathcal{G}_{n+1}(2k-1) - (k-1)2^{-n} \subset \mathcal{G}_n(k) \subset 2\mathcal{G}_{n+1}(2k) - k2^{-n} \tag{16}$$

Sketch of Proof: We only prove the result for $x \notin \mathbb{D}$, since the dyadic rationals are self-similar and since $\mathbb{D} \subset \mathcal{G}$. If $b_k^n = b_1 b_2 \cdots b_n$ is the terminating binary expansion of $(k-1)2^{-n}$, every value in $[(k-1)2^{-n}, k2^{-n}]$ has a binary expansion $b_k^n a$ for some $a \in \{0,1\}^\infty$, where $b_n = 1$ if and only if $(k-1)$ is odd. Similarly, and since $(2k-1)$ is always odd, every value in $[(2k-1)2^{-n-1}, k2^{-n}]$ has a binary expansion $b_k^n 1a'$ for some $a' \in \{0,1\}^\infty$. Assume that $a' = a$. Then, by Lemmas 3 and 4, $W_\infty^{b_k^n a} \preccurlyeq W_\infty^{b_k^n 1a}$ for all a. Hence, if $f(b_k^n a) \in \mathcal{G}_n(k)$, then $f(b_k^n 1a) \in \mathcal{G}_{n+1}(2k)$. The proof follows by showing that $2f(b_k^n 1a) - f(b_{k+1}^n) = f(b_k^n a)$. For the BEC, the proof follows from the fact that by Lemmas 3 and 4, $W_\infty^{b_k^n 0a} \preccurlyeq W_\infty^{b_k^n a}$ for all a. ■

In other words, at least for the BEC, $\mathcal{G}$ is composed of two similar copies of itself (see Fig. 1). Along the same lines, the quasi self-similarity of $\mathcal{B}$ can be shown.

Example 4. By careful computations we obtain $\vartheta(1/6) \approx 0.214$, $\vartheta(1/3) \approx 0.382$, and $\vartheta(2/3) \approx 0.618$. Indeed, if we consider $1/3$ in $\mathcal{G}$, then $1/6$ and $2/3$ are the corresponding values in $\mathcal{G}_1(1)$ and $\mathcal{G}_1(2)$. Since $\vartheta(1/6) < \vartheta(1/3) < \vartheta(2/3)$, for the BEC we have the inclusion indicated in Proposition 6.

IV. Discussion & Outlook

That polar codes satisfy fractal properties has long been suspected: Every nontrivial, partly polarized channel $W_{2^n}^{b^n}$ gives rise, by further polarization, to both perfect and useless channels, regardless how close $I(W_{2^n}^{b^n})$ is to zero or one. This fact is reflected in our Propositions 3 and 4, which state that the good channels are dense in the unit interval (and so are the bad channels for BECs): A partial polarization with sequence b^n corresponds to an interval with dyadic endpoints, and denseness implies that in this interval there will be both perfect and useless channels. Proposition 6, claiming the self-similarity of the sets of good and bad channels, goes one step further and gives these sets structure: If a channel polarized according to the sequence $b^n a$ is good, then so is the channel polarized according to $b^n 1a$. Proposition 2 is also of interest in this context: In [14, Prop. 3], we prove that the thresholds $\vartheta(x)$ are symmetric, in the sense that $\vartheta(1-x) = 1-\vartheta(x)$, a fact that is also visible in Fig. 1.

An obvious extension of our work should deal with the fractal properties of non-binary polar codes. If q is a prime number, then every invertible $\ell \times \ell$ matrix with entries from $\{0, \dots, q-1\}$ is polarizing, unless it is upper-triangular [11, Thm. 5.2]. The n-fold Kronecker product of one of these matrices generates ℓ^n channels. It should be easily possible to design a function mapping $\{0, \dots, \ell-1\}^\infty$ to $[0, 1]$ (cf. (2)), admitting an analysis similar to the one presented in this paper. Since choosing appropriate polarization matrices for non-binary alphabets is not trivial, we propose to evaluate choices based on the properties of the corresponding *polar fractal* (see Fig. 1). This would, in addition to error probabilities or polarization rates, present another objective for the design of non-binary polar codes.

Whether binary or not, it is presently not clear if our infinite-blocklength results can be carried over to practically relevant finite-length codes. If this was the case, a possible application of our results would be code construction, which requires knowledge about the structure of the set of good channels. Future work shall investigate this issue.

Acknowledgments

The author thanks Emmanuel Abbe, Princeton University, and Hamed Hassani, ETH Zurich, for fruitful discussions and suggesting material.

References

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[2] J. M. Renes, D. Sutter, and S. H. Hassani, "Alignment of polarized sets," in *Proc. IEEE Int. Sym. on Information Theory Proceedings (ISIT)*, Jul. 2015, pp. 2446–2450, extended version available: `arXiv:1411.7925 [cs.IT]`.

[3] S. Kahraman, E. Viterbo, and M. E. Çelebi, "Folded tree maximum-likelihood decoder for Kronecker product-based codes," in *Proc. Allerton Conf.*, Oct. 2013, pp. 629–636.

[4] E. Arıkan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447–449, Jul. 2008.

[5] E. Abbe, "personal communication," Nov. 2011.

[6] M. Taylor, *Measure Theory and Integration*, ser. Graduate studies in mathematics. American Mathematical Soc., 2006.

[7] S. Haghighatshoar and E. Abbe, "Polarization of the Rényi information dimension for single and multi terminal analog compression," in *Proc. IEEE Int. Sym. on Information Theory Proceedings (ISIT)*, Jul. 2013, pp. 779–783.

[8] E. Abbe and Y. Wigderson, "High-girth matrices and polarization," Jan. 2015, `arXiv:1501.06528 [cs.IT]`.

[9] R. Nasser, "Ergodic theory meets polarization. I: An ergodic theory for binary operations," Feb. 2015, `arXiv:1406.2943v4 [cs.IT]`.

[10] ——, "Ergodic theory meets polarization. II: A foundation of polarization theory," Feb. 2015, `arXiv:1406.2949v4 [cs.IT]`.

[11] E. Şaşoğlu, "Polarization and polar codes," *Foundations and Trends® in Communications and Information Theory*, vol. 8, no. 4, pp. 259–381, 2011.

[12] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2009.

[13] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[14] B. C. Geiger, "The fractality of polar and Reed-Muller codes," Jun. 2015, in preparation: `arXiv:1506.05231 [cs:IT]`.

[15] S. H. Hassani, K. Alishahi, and R. L. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5875–5898, Oct. 2014.

[16] S. H. Hassani, S. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Sep. 2009, pp. 16–21.

[17] M. El-Khamy, H. Mahdavifar, G. Feygin, J. Lee, and I. Kang, "Relaxed polar codes," Jan. 2015, `arXiv:1501.06091 [cs.IT]`.

[18] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 3rd ed. Chichester: John Wiley & Sons, 2014.